REMARKS

In view of the foregoing amendments and following remarks responsive to the Non-Final Office Action of July 26, 2006, Applicants respectfully request favorable reconsideration of this application.

The Present Invention

The present invention is a method and apparatus for teaching a speech recognition system or a text-to-speech system the proper pronunciations of letters or letter groups within particular character strings, such as words or names (hereinafter strings). Specifically, the particular pronunciation of a letter or letter group (hereinafter letter group) in any given character string can depend on many different factors, including the particular language, the particular word within which it appears, the particular usage of that word (e.g., noun or verb), the particular speaker, etc. The present invention involves a scheme by which a user can enter a character string using a graphical user interface (hereinafter GUI) and then teach the system (e.g., the software) how to pronounce various letter groups in the string. More particularly, the user selects a particular letter group and the software GUI presents the user with a plurality of words containing that letter group. The user can then select the word in the list in which the pronunciation of that letter group most closely matches the correct or desired pronunciation of that letter group in the string. The system also provides similar GUI for allowing the user to change syllable breaks and/or accent within the string.

The Blackmer Reference

Blackmer does not concern a <u>person</u> teaching <u>software</u> the pronunciation of letter groups or words. Rather, Blackmer pertains to <u>software</u> for teaching <u>a person</u> the correct pronunciation of words in a given language. In the latest Office Action, the Office essentially conceded that it cited Blackmer essentially merely for its general teaching of "a user interface for the exchange of pronunciation information" Office Action, page 4, last paragraph.

The Kuhn Reference

Kuhn discloses a computer program for teaching software the pronunciation of letter groups or words. However, the process is very different than the process disclosed in the present application. Kuhn's process is completely automated and involves no human interaction, which is a cornerstone of the present invention. There are no lists of words having different possible pronunciations of a character string.

Discussion

The Office has maintained all of the rejections asserted in the two previous Office Actions, but has added one explanatory paragraph further explaining the basis of the rejections.

All claims, claims 1-23, stand rejected under 35 U.S.C. 103(a) as unpatentable over Blackmer in view of Kuhn.

Applicant maintains that, not only do Blackmer and Kuhn not make the present claims unpatentable, but that neither Blackmer nor Kuhn contains any teachings particularly relevant to the present invention as claimed. Specifically, the present invention is a technique for training a speech synthesis computer system the correct pronunciation of words. Blackmer does not pertain to this topic. Blackmer teaches a system for using a computer to teach a person the correct pronunciation of words, which is basically the inverse of the present invention.

Blackmer Does Not Disclose That For Which It Has Been Cited

The Office's reliance on Blackmer is paradoxical because, on the one hand, the Office essentially conceded that it cites Blackmer only for its teaching of a graphical user interface in the context of speech recognition, but, on the other hand, in the detailed explanation of the rejection, asserts that Blackmer teaches all but one of the recited elements of all 23 claims. Particularly, the Office cites Kuhn for one and only one teaching, namely, "training the recognition system for pronunciation".

However, despite the minimal use of Kuhn in the rejections, the reliance on Kuhn

for this teaching makes the rejections especially paradoxical because that is exactly the fundamental subject matter of all of the claims, i.e., a system for training a recognition system in pronunciation.

It is not understood how the Office can conclude that a reference that admittedly does not teach "a system for training the recognition system for pronunciation" could teach all of the elements of a claim directed to training a recognition system for pronunciation.

The following is a detailed traversal of the rejections of the independent claims 1, 15, and 23.

The Office asserted that Blackmer "teaches a computer allowing the user to set a pronunciation of a string of characters (figure 1 subblock 19)", as recited in the preamble (last two lines of page 2 Office Action).

This is utterly inconsistent with the Office's concession six lines later on line 6 of page 3 that Blackmer "does not explicitly teach training the recognition system for pronunciations".

In any event, Blackmer clearly does <u>not</u> teach a system in which a user sets the pronunciation of words. As previously explained, Blackmer pertains to the inverse process, i.e., the computer teaching a person the pronunciation of words. Blackmer starts from the presumption that the software already knows the pronunciations of the words.

Furthermore, the Office's reference to figure 1, subblock 19 of Blackmer is even more confusing. Figure 1 of Blackmer is a block diagram in which block 19 is a graphical user interface. The graphical user interface 19 does not pertain to allowing a user to set a pronunciation of a string of characters since, as previously explained, that is not the subject matter of Blackmer.

The Office further asserted that Blackmer teaches the first two steps of claim 1 of "allowing the user to select a set of one or more characters in a particular one of the

-10-

strings of characters" and "retrieving from a database... a plurality of samples of words
or parts of words representing different possible pronunciations of the selected
character set and displaying the received samples" at column 19, lines 1-34[1].

This is not true. Column 19, lines 1-34 describe how to navigate through various
GUIs of a lesson on the difference in pronunciation between the letters "p" and "b". It
mentions that icons 644 and 646 lead to lessons on pronouncing "p" and "b",
respectively. However, it does not describe the lessons (not that it would matter if it did
since the present invention does not concern teaching people how to pronounce words,
but teaching the computer how to pronounce words).

There is no mention in Blackmer of different possible pronunciations of the
letters "p" or "b". In fact, the obvious premise of Blackmer is that each of the letters is
pronounced the same way in all of the displayed words. There is no "retrieving a
plurality of syllables... representing different possible pronunciations".

Accordingly, Blackmer teaches none of the things that the Office has asserted
that it teaches.

The Proposed Combination of Blackmer and Kuhn Is Improper

The Office then asserted that Kuhn "teaches incorporating the pronunciation
aspect into existing systems" and, therefore, it would have been obvious to modify the
teachings of Blackmer with training an existing recognition system because it would
advantageously provide useful feedback to the user with respect to pronunciation
accuracy".

While the Office's conclusion is flawed, Applicant certainly does not dispute that
it would be obvious that Blackmer's system would need to be trained how to pronounce

---

1 Applicant has slightly changed the language of the claims to (1) change the term "samples of words" to "sample
words" to simplify the claim language and (2) to eliminate the recital of "parts of words", which was extraneous
language previously included for purposes of noting that the pronunciation in question most likely would be
represented by only part of the sample word, rather than the whole word, but which language now seems, at a
minimum, superfluous and, at worst, subject to undesired interpretations.

words before it could, in turn, be used to train persons how to pronounce words.

However, having said that, it is not seen how a first reference teaching a GUI for interfacing with a user to achieve a task that is entirely inapposite to the task that the claimed invention performs and a second reference that, admittedly, achieves the same ultimate task as the claimed invention, but using a completely different technique that does not even employ a GUI (because it uses an automated process involving no human interaction), could possibly suggest the invention.

Since Blackmer does not address the issue of training a recognition system at all, the only combination that could possibly be suggested is using Kuhn's system-training technique to train Blackmer's user-training system. This, of course, is not even really a modification of Blackmer since Blackmer does not discuss how the system was trained in the first place.

In any event, while Kuhn does, in fact, disclose a system for training a recognizer how to pronounce words, it is a completely different system than the one claimed in the present application that does not even have a GUI. Therefore, the combination could not possibly teach the elements of the present invention. The Office does not contend otherwise. The Office uses Kuhn solely for suggesting that Blackmer's system must first be trained before it can be put into use, but relies on Blackmer as teaching all of the specific elements of the claims.

Certainly there is nothing suggesting to a skilled artisan any combination of aspects of Blackmer's person-training technique with aspects of Kuhn's system-training technique. These tasks are inapposite to each other. Applicant reminds the Office of its previous analogy: The present situation is no different than if Kuhn disclosed how to fabricate a tennis racquet and Blackmer disclosed how to play tennis. Two such references contribute to the playing of a game of tennis, but one reference does not contain anything that can be substituted into the other reference.

Kuhn Does Not Teach That Which The Office Asserts

Turning now to more specific consideration of the Kuhn reference, it teaches a technique using decision trees for generating an ordered list of possible pronunciations of a spelled word (from more likely correct to less likely pronunciations). With reference to the flowchart of figure 1, in a first stage, a word is input in block 14 to a dynamic programming phoneme sequence generator 16 that comprises a letter-only decision tree 10 to generate a list of pronunciations 18 representing possible pronunciation candidates of the spelled word. The sequence generator 16 constructs one or more pronunciation hypotheses that are stored in a list 18, each associated with a numerical score as to the probability that it is correct.

In a second stage, the same word is input to a different, mixed tree score estimator 20 that re-scores each of the pronunciations in the list 18 based on mixed-tree questions. The mixed-tree score estimator 20 uses a set of mixed-decision trees 12 to assess the viability of each pronunciation in list 18. The score estimator 20 works by sequentially examining each letter in the input sequence along with the phonemes assigned to each letter by sequence generator 16.

A selector module 24 accesses the list 22 to retrieve one or more of pronunciations of the list. Typically, selector 24 retrieves the pronunciation with the highest score and provides this as the output pronunciation 26.

As previously noted, the present Office Action appears to be a verbatim copy of the previous Office Action, except for the addition of a single paragraph. That paragraph is enlightening as to the Office's rationale and, particularly, as to the portions of Kuhn upon which the Office is relying. That paragraph is quoted below for ease of reference:

> The combination of Blackmer et al (5393236) in view of Kuhn et al (6016471) teaches storing the pronunciation data by a computing system for pronouncing the string of characters (Kuhn et al (6016471), teaches generating a n-best list of pronunciations - figure 1, label 22, wherein the pronunciations from the list can

-13-

be stored and used to create pronunciation dictionaries, especially during a training phase – column 5 lines 10-15); furthermore, combination of Blackmer et al (5393236) in view of Kuhn et al (6016471) teaches user selection/feedback with respect to desired pronunciations (Kuhn et al (6016471), column 5 lines 1-10 - wherein a lexicographer can use the feedback to design their own pronunciation database; and lines 20-25).

It is clear that the Office considers column 5, lines 1-10, lines 10-15, and lines 20-25 to be significant. Accordingly, Applicant will herein address of those portions of Kuhn.

Column 5, lines 1-25 are reproduced below for convenient reference with underlining and parentheticals indicating the specific portions referred to by the Office.

This situation might be encountered where a previously developed pronunciation dictionary is available. In such case the mixed-tree score estimator 20, with its associated mixed trees 12, may be used to score the entries in the pronunciation dictionary, identifying those having low scores, thereby flagging suspicious pronunciations in the dictionary being constructed. Such a system may, for example, be incorporated into a lexicographer's productivity tool. (lines 1-10)

The output pronunciation or pronunciations selected from list 22 can be used to form pronunciation dictionaries for both speech recognition and speech synthesis applications. In the speech recognition context, the pronunciation dictionary may be used during the recognizer training phase by supplying pronunciations for words that are not already found in the recognizer lexicon. (lines 10-15) In the synthesis context the pronunciation dictionaries may be used to generate phoneme sounds for concatenated playback. The system may be used, for example, to augment the features of an E-mail reader or other text-to-speech application. The mixed-tree scoring system of the invention can be used in a variety of applications where a single one or list of possible pronunciations is desired. For example, in a dynamic on-line dictionary the user types a word and the system provides a list of possible pronunciations, in order of probability. (Lines 20-25)

Lines 1-10 discuss a situation in which a previously developed pronunciation dictionary is available. It states that, in such cases, the mixed-tree score estimator 20 may be used to score the entries in the pronunciation dictionary and may be incorporated into a lexicographer's productivity tool.

This merely discloses a possible application of Kuhn's fully automated system and not, as the Office would have it, a modification of Kuhn employing user feedback in the process of training the system. The Office's assertion that this suggests that Kuhn teaches user selection/feedback with respect to the desired pronunciations is erroneous. Furthermore, even if it did, it does not teach anything resembling the interface technique of the present invention. Particularly, Kuhn's vague statement that the system "may, for example, be incorporated into a lexicographer's productivity tool" certainly cannot in any rational way be interpreted as suggesting anything other than a potential application of the fully automated technique disclosed throughout the entire specification of Kuhn. It is not even clear what is meant by "lexicographer's productivity tool". There is nothing to suggest that any such "tool" might involve user interaction. Even if it did involve user interaction, there is absolutely no disclosure as to what that interaction might be. Accordingly, this section of Kuhn could not possibly suggest to a person of ordinary skill in this art the technique of the present invention.

Turning to lines 10-15, it merely states the obvious, i.e., that the output pronunciations generated by Kuhn can be used to form pronunciation dictionaries for speech recognition or speech synthesis applications, which may be used during recognizer training. This section of Kuhn says nothing other than that the technique disclosed in Kuhn can be used to train a speech recognizer the pronunciation of words. Applicant has never disputed that Kuhn teaches a technique for training a speech synthesis system the pronunciation of words. It is just a completely different technique than the present invention.

Finally, lines 20-25 merely disclose another application for the disclosed technique, namely, a dynamic online dictionary in which a user types a word and the system provides a list of possible pronunciations in order of probability.

Once again, this is nothing more than a potential application of Kuhn. It does not propose an alteration to the technique, which, as always, is fully automated and does

not involve user interaction. In fact, this portion of Kuhn is particularly inapposite because, like Blackmer, it discusses the use of the system to train humans, whereas the present invention concerns humans training the system.

The Prior Art of Record Is Fundamentally Different From the Present Invention Such That No Reasonable Case for Obviousness Exists

The present invention differs from Kuhn (and Blackmer) in a fundamental way. The system of the present invention is based on the fundamental premise that it is easier for a human to interact with and train the system by using sample, real words containing the character set in question and picking the sample word whose pronunciation of that character set is the same pronunciation in the word being trained. This technique is much simpler than prior art phonetic spelling techniques.

Neither Blackmer nor Kuhn discloses anything even remotely resembling this. Thus, no combination of these two references could teach such to a person of ordinary skill in the related arts.

The Independent Claims Specifically Distinguish Over the Prior Art

Turning to claim 1, neither Blackmer nor Kuhn nor any combination thereof discloses "retrieving from a database accessible by the computer a plurality of sample words representing different possible pronunciations of the selected character set and displaying the retrieved sample", "allowing the user to select one of the displayed sample words", or "updating the pronunciation data corresponding to the particular string of characters in accordance with a pronunciation of the selected character set in the sample word selected by the user".

The other two independent claims 15 and 23 distinguish over the prior art of record for the same basic reasons. Specifically, claim 15 contains essentially identical language to the language of claim 1 discussed above. Claim 23 includes similar language also.

The Dependent Claims Add Further Distinguishing Limitations

The dependent claims contain even further distinguishing features. For instance, claim 2 adds "generating a pronunciation of the character string using the pronunciation represented by the sample selected by the user as the pronunciation for the selected character set, and audibly outputting the generated pronunciation". Blackmer cannot disclose this since there is no pronunciation that is "selected by the user".

Claim 4 adds the step of "allowing the user to select a second of the displayed samples and storing second pronunciation data comprising the string of characters with the selected character set being assigned a pronunciation represented by the second sample selected by the user". This is a similar step to the above-discussed step in claim 1. This claim just adds the feature of allowing the user to select a second, alternate pronunciation of the word/character string. It distinguishes over the prior art for all of the same reasons as discussed above in connection with similar step of claim 1.

Claim 5 depends from claim 4 and adds the step of, during a text-to-speech process of generating audible output of a text file containing a string of characters, selecting one of the first or second pronunciation data. The Office asserted that this is found in Blackmer, col. 19, lines 28-39, which were discussed above in connection with claim 1. Since Blackmer does not discuss how the computer selects pronunciation of the words, it cannot possibly disclose this feature.

Claim 6 depends from claim 5 and further adds the limitation of "associating the first and second pronunciation data with first and second objects, respectively, and selecting one of the first and second objects, and during the step of selecting one of the first and second pronunciation data comprises selecting the pronunciation data associated with the selected object". This claim builds on the feature recited in claims 4 and 5 where a letter group in a single word may have a different pronunciation depending on context (i.e., the object). Claims 7 and 8 continue to build on this

-17-

concept. Claim 7 recites that the particular pronunciation selected by the software is selected based on the pronunciation of the particular user as determined during a speech recognition process. Claim 8 is very similar to claim 6 except depending from claim 7. The Office asserted that all of this is found in Blackmer, col. 9, line 36 - col. 10, line 45. However, as noted above in the discussion of claim 1, neither this nor any portion of Blackmer discloses anything about how the software determines the pronunciation. Blackmer only discusses how the software teaches the user how to pronounce words. Blackmer starts from the presumption that the software has this information and does not discuss where this information came from. Accordingly, it could not possibly teach any of the features recited in claims 6-8, which concern how the system learns the correct pronunciation.

Claim 9 pertains to the feature discussed on page 8, line 15-18 of the specification of allowing the user to alter the syllable breakdown of a word from the default breakdown provided by the computer. The Office asserted that this is taught in column 22, lines 20-25 of Blackmer. However, that portion of the specification merely notes that the plurality of words shown in the GUI represented by Fig. 9C has "a like number of syllables, and, furthermore, the same syllable in each of the words is stressed". This is utterly irrelevant to what is claimed in claim 9.

Claim 10 recites "allowing the user to identify a part of the character string to associate with an accent, and wherein the step of storing said first pronunciation data comprises storing data representing the identified accent". Claim 10 recites the feature disclosed on page 8, line 19 through page 9, line 6 of the specification wherein the user can change the syllabic accentuation in the word as desired. The Office asserted that this is disclosed in Blackmer in column 22, lines 18-31 (the same portion referred to in connection with claim 9 discussed immediately above). However, as noted above in connection with claim 9, this portion of Blackmer has nothing to do with teaching the

-18-

computer how to pronounce the word, but instead relates to the computer teaching the user how to pronounce the word.

Claims 16, 17, 18, 19, 20, and 21 depend directly or indirectly from claim 15 and contain limitations similar to those discussed above in connection with dependent claims 2, 3, 4, 5, 6, and 7, respectively. Hence, claims 16, 17, 18, 19, 20 and 21 further distinguish over the prior art for the reasons discussed with respect to claims 2, 3, 4, 5, 6, and 7.

Conclusion

In view of the foregoing amendments and remarks, this application is now in condition for allowance. Applicant respectfully requests the Office to issue a Notice of Allowance at the earliest possible date. The Examiner is invited to contact Applicant's undersigned counsel by telephone call in order to further the prosecution of this case in any way.

Respectfully submitted,

Dated: November 22, 2006

  /Theodore Naccarella/
Theodore Naccarella
Registration No. 32,023

Synnestvedt & Lechner LLP
2600 Aramark Tower
1101 Market Street
Philadelphia, PA 19107
Telephone: (215) 923-4466
Facsimile: (215) 923-218

Attorneys for Applicant

TXN:pmf

S:\L\LUCENT TECHNOLOGIES, INC\PATENTS\P23141 USA\PTO\RESPONSE TO NON-FINAL OX.DOC